How textual data, neighbourhood and amenities have influence in Airbnb prices in London

Elisa Chimenton

Abstract— This project is looking at Airbnb listings and reviews dataset to find hidden adding value features from texts that can be valuable to visitors when going to book their accomodation and also find adding value features to hosts to help to have better balance between occupancy and high earnings. Correlations, text processing and difference in means are some of methodologies used to attempt to answer commons problems in tourist industry.

1 INTRODUCTION

Airbnb is a house-sharing service that give opportunity to travellers to find affordable accommodation around the world. It is known that location is one of the main factor impacting price but probably there are more factors influencing it. London is one of the most famous touristic place in Europe and people participate actively in sharing services. Greater London is divided into 32 boroughs and the city of London – a separate county but in this study it will appear as a borough. The main objective of this analysis is provide an overview about Airbnb London listings providing average price per person in different boroughs and identifying value-adding features that would help London visitors find the best options to stay.

2 DATA, QUESTIONS AND PLAN

2.1 Data source

The dataset used for this project comes from Inside Airbnb, and it is include Airbnb listings and reviews from London but is possible to find from many cities around the world. The dataset was scraped on 05 November 2019 and contains information on all London Airbnb listings that were live on the site on that date.

It has some limitations, it includes the advertised price (sometimes called the 'sticker' price). The sticker price is the overall nightly price that is advertised to potential guests, rather than the actual average amount paid per night by previous guests. The advertised prices can be set to any amount by the host, and hosts that are less experienced with Airbnb will often set these to very low (e.g. £0) or very high (e.g. £10,000) amounts but even if this was not real price can show good insights.

2.2 Research Questions

Data analytics problems in tourism and hospitality industry related to Airbnb can be focused on tourists, hosts or stakeholders. Home sharing platform allows home-owners known as hosts to set their own prices for their listings. Price is a key component for hosts but also for travellers when planning trips, specially to expensive cities such as London. What motivated this work was the real necessity to help tourist identify fair prices in different covered topics during search process because normally when searching for accommodations in online platform, visitors have words in reviews, title descriptions, amenities and pictures to help they decide better place to stay but the latter is out of scope of this analysis. Now taking in consideration that normally people do not know well place they are going to visit, seems important analyse host’s text and also guests reviews to find what aggregate or not value to house’s price. Understanding better house-sharing service would be possible provide pleasant experience to tourists with fair prices and also balance high earnings with occupancy to hosts.

With this objective in mind, some questions will be asked to help define better scope of this analysis:

- How much is the average price per person in London at different neighbourhoods?
- Is the average price correlated with ratio of positive comments in this location?

- How much extra will travellers pay on accommodation advertised as luxury, having garden, spacious, modern, rooftop, etc.?
- Which amenities are common in listings?

Responding these questions, we aim to uncover text patterns used by hosts to increase their house prices with or without real value and also find features that add real value to visitors, reaching conclusion that can help tourist find fair prices to stay and also hosts to improve their earnings.

2.3 Analysis strategy

In order to have a better understand about the target variable some graphs was plotted as figure 1. This figure shows how the price are distributed geographically, the result showed was already expected, accommodations closer to city centre are more expensive than the rest, initial analysis helps understand better the data, to this analysis some steps was followed:

1- Load data
2- Clean data – reshape, handling missing values, detecting and filtering outliers
3- Data understanding
4- Engineers new features, check correlations
5- Modelling

First objective will be calculate how price differs in relation to number of people a house accommodates, then calculate average price per person per borough for different room types and compare among them.

Second objective is used text processing techniques to identify if reviews with positive comments is in line with price in this location. For example, check if more expensive areas receive best reviews or identify if some places are over or under priced according to their location.

Then, the third part will observe what factors add-value to Airbnb accommodations by analysing listings description titles at this part, the reviews dataset will be divide in two, separating features with and without keywords as luxurious, spacious, modern. Thus compare distribution of prices in this two datasets and construct a confidence interval for the difference in means.

The last part will analyse the most common amenities in listings.

3 FINDINGS AND REFLECTIONS

3.1 Overview neighbourhood

The first analyse was made to understand the data was check how price is distributed geographically, what was already expected, accommodations closer to city centre are more expensive than the rest.

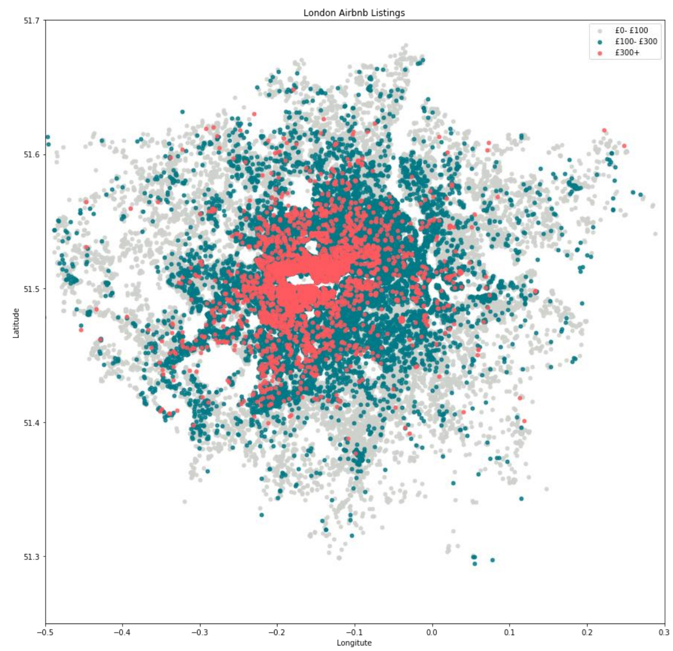


Figure 1: Overview prices distribution in London geographically.

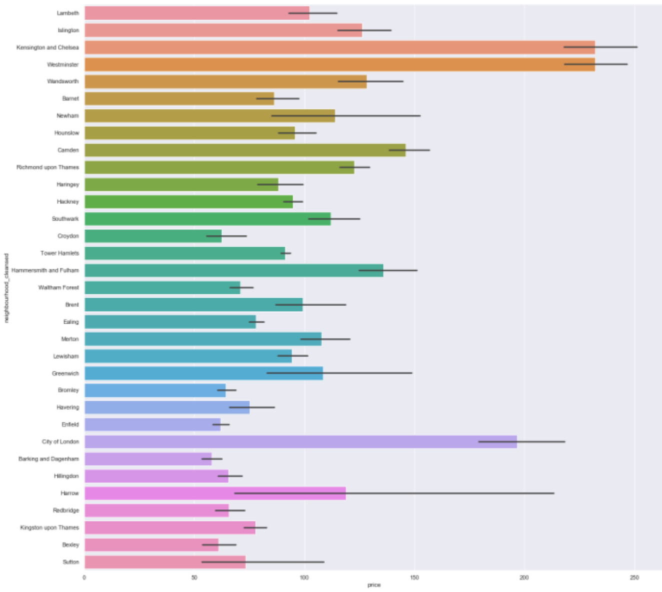


Figure 2: Histogram average price per boroughs

3.2 Compare average price per room types

As showed in the map above the central areas are most expensive, but how much would be this difference per person when renting a entire home or shared room. The results showed when filtered by types of house, shared rooms is much more affordable to tourists, would be possible save around £40 in this kind of accommodation compared to entire homes or apartments and this difference cover all boroughs. City of London is the unique borough that price can go over £100 in shared room, thus in this borough price is much higher having average of £175 per person in entire homes. This difference

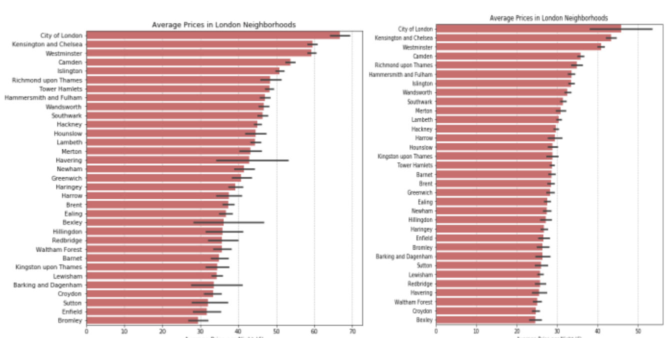


Figure 3: Averages prices in London per types of accommodations.

3.3 Compare the percentage of guests review with positive comments about location

Houses prices in London varies significantly related to location and it is expected that Airbnb also follow similar trends but one of the most important things about success of sharing economy services are reviews, seems interesting to analyse what guests post in these reviews looking for positive statements about location and evaluate if price are in line with the price difference showed in figure 2 above.

Looking at figure 4 is possible see that average price per person is strong correlated to price having only few boroughs that seems over or under priced given their location.

This analyse seems support the theory of market price is measured by location in most of cases. According to Gutierrez et al. (2017), Airbnb listings are mainly located near tourist attractions, this can be a reason for City of London, Kensington and Chelsea be overpriced however more analyses should be made to understand the in depth reasons of these outliers. Kingston upon Thames and Greenwich seems be much cheaper in relation positive location review analysed and Westminster shows that Airbnb market price for this location is fair to the market.

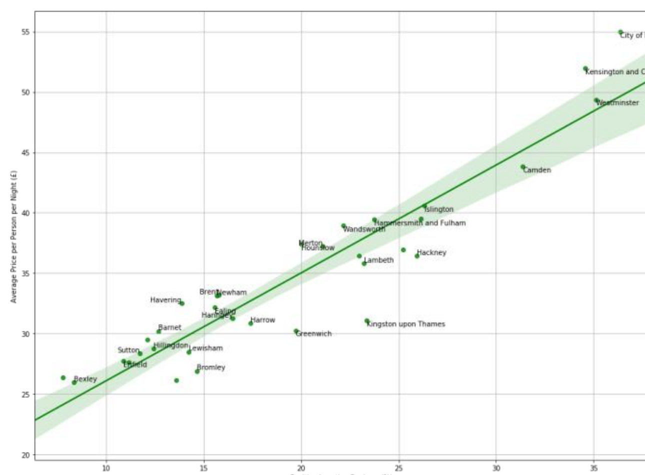


Figure 4: Correlation between average price per person per night and positive location reviews.

3.4 Identifying adding-value features from listings description titles

Analysing listings dataset is possible to see that there are several aspects to differentiate entire house or apartment prices, at this part was observed listings' textual descriptions to identify features demonstrated by keywords that may add value to properties. For this was created word Clouds with descriptions of apartments in all neighbourhoods, the first word Clouds created appeared many words as double, bedroom, central, private room that is not relevant to answer this question but it means that most of shared rooms describe their bedroom with these characteristics, then these words was excluded and the result was word Clouds figure 4 below.

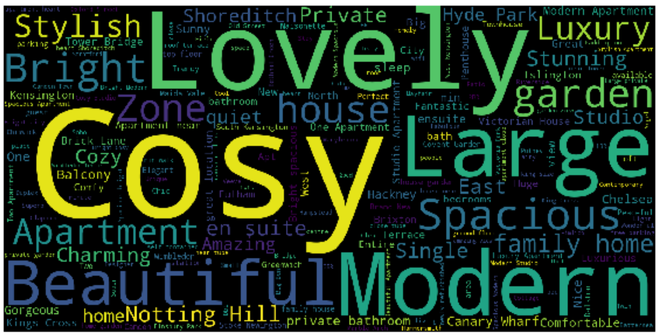
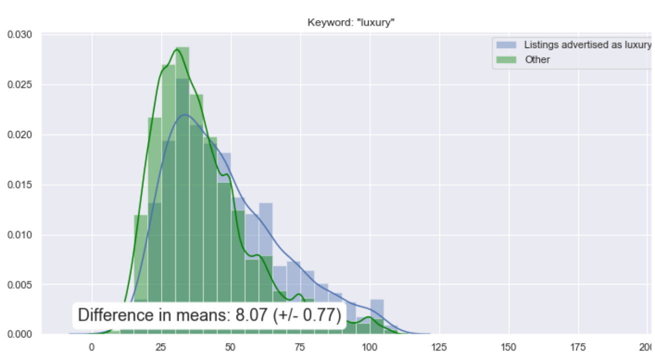


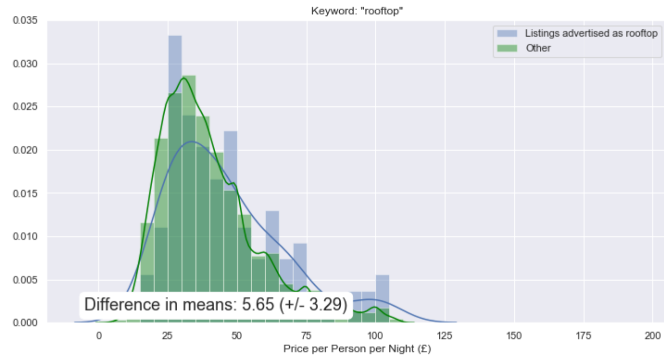
Figure 5: Word Clouds from description tiles

Entire houses or apartments described as luxury are on average £8 more expensive than other Airbnb accommodation without this description.

Keywords as Cosy, Large, Lovely, Spacious, Modern seems do not increase any value, they may are used as a way to attract people when do not other features to describe.

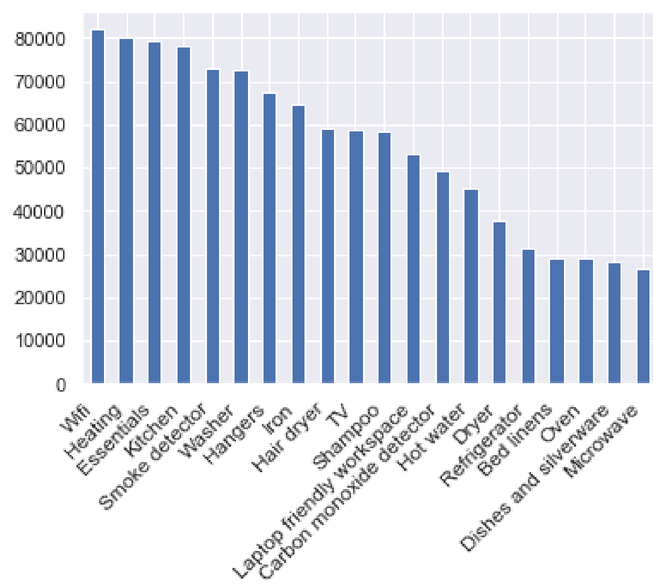
Another keyword that added value to title was Rooftop increasing on average £5.





3.5 Amenities

Looking at list of all amenities included at listings dataset can be deducted that some amenities can be more important than others (e.g. Dryer is more likely to increase price than a fax machine), and some are likely to be fairly uncommon (e.g. 'Electric profiling bed') but also is important hosts know which is the most common amenities because they can miss the opportunity to rent their homes without it, for example Wi-Fi is expected for guests and seems be present at most of listings and it probably doesn't increase price but hosts can lose opportunity to rent their place without it. But in cases as iron hair dryer, shampoo that is not expensive items if hosts do not have it, they can think about to buy it and increase their rent and guests experience that can result in good reviews.



References:

[1] Mingming Cheng, Xin Jin,. What do Airbnb users care about? An analysis of online review comments, *International Journal of Hospitality Management*, Volume 76, Part A,2019,Pages 58-70.

[2] Chuhan (Renee) Wang, Miyoung Jeong. What makes you choose Airbnb again? An examination of users' perceptions toward the website and their stay.*International Journal of Hospitality Management,* Volume 74,2018,Pages 162-170, ISSN 0278-4319.

[3] Manojit Chattopadhyay, Subrata Kumar Mitra. Do airbnb host listing attributes influence room pricing homogenously? *International Journal of Hospitality Management*,Volume 81,2019,Pages 54-64.

[4] Abdelaziz Lawani, Michael R. Reed, Tyler Mark, Yuqing Zheng, Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, Volume 75, 2019,Pages 22-34.

[5] Feifei Xu, Mingxing Hu, Liqing La, Jialing Wang & Chao Huang The influence of neighbourhood environment on Airbnb: a geographically weighed regression analysis, *Tourism Geographies*, 2019.